Overview of the MPEG-7 Standard

Shih-Fu Chang, *Senior Member, IEEE*, Thomas Sikora, *Senior Member, IEEE*, and Atul Puri, *Member, IEEE*

***Abstract*—MPEG-7, formally known as Multimedia Content Description Interface, includes standardized tools (descriptors, description schemes, and language) enabling structural, detailed descriptions of audio–visual information at different granularity levels (region, image, video segment, collection) and in different areas (content description, management, organization, navigation, and user interaction). It aims to support and facilitate a wide range of applications, such as media portals, content broadcasting, and ubiquitous multimedia. In this paper, we present a high-level overview of the MPEG-7 standard. We first discuss the scope, basic terminology, and potential applications. Next, we discuss the constituent components. Then, we compare the relationship with other standards to highlight its capabilities. Some parts of the standard are also covered in depth in this Special Issue.**

***Index Terms*—Content description, interoperability, ISO/IEC Standard, metadata, MPEG-7, multimedia indexing.**

I. Introduction

MPEG-7, formally known as Multimedia Content Description Interface, is the next ISO/IEC standard under development by MPEG, following the successful development of the MPEG-1, MPEG-2, and MPEG-4 standards. While the prior standards focus on coding and representation of audio–visual content, MPEG-7 focuses on description of multimedia content. It addresses content with various modalities including image, video, audio, speech, graphics, and their combinations. MPEG-7 complements the existing MPEG standard suite and aims to be applicable to many existing formats, which include non-MPEG formats and noncompressed formats as well.

Originating in 1998, MPEG-7 aims at becoming an international standard in September 2001. MPEG-7 follows the same development process as those used in prior MPEG standards: definition of context and objective, identification of requirements, call for proposals, evaluation, and convergence to the final specification. The participating groups represent companies and institutions from different sectors, such as software developers, manufacturers, service providers, broadcasters, academics, and libraries.

Motivation for developing the MPEG-7 standard is driven by the clear trends in technology, market, and user needs. Proliferation of tools for audio–visual content creation, digitization, and distribution has brought about an environment in which a large amount of audio–visual content is available. Emerging broadband solutions for the last mile are removing one major barrier preventing users from accessing broadband content. On the user side, the emerging multidevice paradigm is empowering users with unprecedented technologies addressing users' personalized contextual information needs.

In such a ubiquitous, multi-device, information-rich environment, critical needs exist for tools and systems for indexing, searching, filtering, and managing audio–visual content, stored or live. Such solutions are important not only for the end users, but also the providers of content or services. In order to achieve the maximum interoperability and facilitate the creation of innovative applications, MPEG-7 intends to be an interoperable interface, which defines the syntax and semantics of various description tools. Each tool may be designed for specific or generic modalities (e.g., audio, visual, or multimedia), aspects (e.g., media, meta, structural, or semantic), and applications (e.g., search engine, filtering agent, navigation).

In response to the needs of the industries and users, many groups and organizations have initiated active works in defining interoperable frameworks and representations for metadata description. The notable ones related to MPEG-7 include Society of Motion Picture and Television Engineers (SMPTE), European Broadcasting Union (EBU), Dublin Core, Digital Imaging Group (DIG), TV-Anytime, and the Ohio Online Computer Center/Research Libraries Group (OCLC/RLG). These approaches share similar objectives and have developed significant correspondence between their frameworks and elements. However, there is also notable divergence that will require nontrivial harmonization and synchronization.

In this paper, we first discuss the scope of MPEG-7, and distinguishing the normative and nonnormative components of MPEG-7. We will also describe the envisioned architecture for using MPEG-7 in applications. Then, we will briefly introduce different parts of MPEG-7 standard: systems, Description Definintion Language (DDL), audio–visual Descriptor (D), and Description Schemes (DSs). Each topic is also covered in depth in separate papers in this Special Issue on MPEG-7 [1]. Finally, we discuss the relationship between MPEG-7 and other metadata standards, and outstanding challenging issues of the standard.

II. Scope of the Standard

The ultimate goal and objective of MPEG-7 is to provide interoperability among systems and applications used in generation, management, distribution, and consumption of audio–visual content descriptions. Such descriptions of streamed or stored media help users or applications to identify, retrieve, or filter audio–visual information. Examples of applications include broadcast media selection (e.g., personalized

Manuscript received September 2000; revised March 21, 2001.

S.-F. Chang is with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA.

T. Sikora is with the Department of Interactive Media—Human Factors, Heinrich–Hertz Institute for Communication Technology, D-10587, Berlin, Germany.

A. Puri is with AT&T Labs—Research, Red Bank, NJ 07701 USA (e-mail: apuri@research.att.com).

Publisher Item Identifier S 1051-8215(01)04985-0.

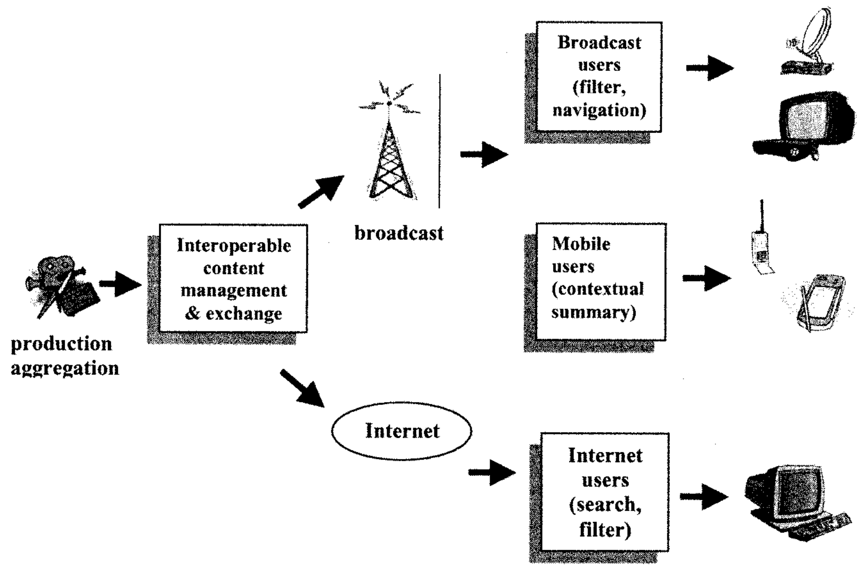

Fig. 1. Role of MPEG-7 in facilitating interoperable services and applications.

radio, TV channels), digital libraries (image catalog, musical dictionary), multimedia directory services (e.g., yellow pages), and multimedia editing (e.g., personalized electronic news service, media authoring). As shown in Fig. 1, the use of MPEG-7 descriptions is expected to result in a flexible and scalable framework for designing services that can be accessed from a variety of terminals such as mobile devices, set top boxes, and personal computers.

Among this diversity of possible applications the MPEG-7 descriptions allow users or applications to perform the following tasks.

- *Multimedia*—Generate a customized program guide or summary of broadcast audio–visual content according to user's preference and use history.
- *Archive*—Generate descriptions of individual elements or the entire collection of audio–visual content, and seamlessly exchange content and corresponding descriptions among content owners, aggregators, and consumers.
- *Adaptation*—Filter and transform the multimedia streams in resource limited environments (e.g., mobile networks and terminals) by matching the user preference, available resources, and content descriptions.
- *Music/Audio*—Pay a few notes on a keyboard and get in return a list of musical pieces containing (or close to) the tune somehow matching the notes.
- *Graphics*—Draw a few lines on a screen and get, in return, a set of images containing similar graphics, logos, or ideograms.
- *Movement*—With a given set of video objects, describe movements and relations between objects and get, in return, a list of animations or video clips fulfilling the described temporal and spatial relations.
- *Scenario*—On a given audio–visual content, describe actions and get in return a list of scenarios (i.e., audio–visual segments) where similar actions take place.

In order to handle such a wide range of applications, MPEG-7 provides several normative elements, including a D, DS, and DDL, and specifications addressing *systems*-level issues. Each of these parts along with the other components of the standard will be discussed further in the next section.

Ds define syntax and semantics of features of audio–visual content. Different levels of abstraction are addressed by MPEG-7. At the low abstraction level, Ds may include shape, motion, texture, color, and camera motion for images/videos, energy, harmonicity, and timbre for audio. At the high abstraction level, Ds may include events, abstract concepts, content genres, etc. Audio and visual Ds represent specific features related to audio and visual content respectively. Generic Ds address generic features.

DSs allow construction of complex descriptions by specifying the structure and semantics of the relationships among the constituent Ds or DSs. For example, the description scheme for a video segment may specify the syntax and semantics of the component elements such as underlying segment decomposition, individual segment attributes (e.g., segment length, textual annotations), and relationships between component segments. As in the case of Ds, DSs can be categorized to audio, visual, or generic. Generic DSs usually represent generic meta information related to all kinds of media (audio, visual, text, graphic, etc).

Next to having Ds or DSs that are derived intrinsically from the content, MPEG-7 also includes Ds and DSs related to creation, production, management, and access of audio–visual content. Such meta-data may include information about the coding scheme used for compression of content (e.g., JPEG, MPEG-2), the overall data size, conditions for accessing the material (e.g., intellectual property rights information and financial information), classification (include parental rating, and content classification into a number of pre-defined categories), and links to other relevant material (the information may help the user speeding up the search).

Fig. 2. Scope of MPEG-7.

The MPEG-7 DDL allows flexible definition of MPEG-7 DSs and Ds based on XML Schema. The current Ds and DSs are application independent. When it is required to describe content for specific domains (e.g., news, films), there is often a need to extend and specialize the generic MPEG-7 tools and use DDL to define specialized or additional tools.

MPEG-7 allows descriptions of audio–visual content at different perceptual and semantic levels. It is expected that low-level features (such as color and structural features) can be automatically extracted in fully automatic ways, whereas high-level features—in particular, those who describe semantic information—need (much) more human interaction or textual annotation by humans or agents.

Fig. 2 depicts the MPEG-7 processing chain to explain the scope of the MPEG-7 standard in a simple way. A typical application scenario involves that MPEG-7 descriptions are extracted (produced) from the content or associated sources. It is important to understand that for Ds and DSs, the MPEG-7 standard does not specify how to extract these descriptions—to leave maximum flexibility to the various applications. Extraction for most parts of the MPEG-7 descriptions is thus not normative. However, as a normative requirement, the representation of these descriptions must conform to the MPEG-7 standard. Compliant MPEG-7 binary or nonbinary descriptions can be accessed, understood, and consumed by applications that are able to decode and process MPEG-7 descriptions. How the MPEG-7 descriptions ought to be used for further processing—i.e., for search and filtering of content—is again not standardized in MPEG-7, to leave maximum flexibility to applications.

It should also be noted that MPEG-7 descriptions may be physically located with the associated audio–visual material in the same data stream or on the same storage system. Alternatively, the descriptions could also be located anywhere else, as long as it is possible to link audio–visual material and their MPEG-7 descriptions efficiently.

III. Components of MPEG-7

The MPEG-7 specification includes a standardized set of Ds and DSs for audio, visual, and multimedia, and a formal language for defining DSs and Ds. MPEG-7 also specifies systems, software, and conformance. Formally, the MPEG-7 standard is referred to as ISO 15938, and is organized into the following parts:

1) ISO/IEC 15 938-1: MPEG-7 Systems
2) ISO/IEC 15 938-2: MPEG-7 Description Definition Language
3) ISO/IEC 15 938-3: MPEG-7 Visual
4) ISO/IEC 15 938-4: MPEG-7 Audio
5) ISO/IEC 15 938-5: MPEG-7 Multimedia DSs
6) ISO/IEC 15 938-6: MPEG-7 Reference Software
7) ISO/IEC 15 938-7: MPEG-7 Conformance

Part 1), MPEG-7 Systems, specifies system level functionalities, such as preparation of MPEG-7 descriptions for efficient transport/storage, synchronization of content and descriptions, and development of conformant decoders. Fig. 3 shows a high-level architecture of a terminal that uses MPEG-7 descriptions, and is referred to as an MPEG-7 terminal. The MPEG-7 data is obtained from transport (e.g., MPEG-2, IP) or storage (e.g., MP4 file) and handed over to the delivery layer that allows extraction of elementary streams by undoing the transport/storage specific framing and multiplexing, and retains timing information needed for synchronization. The elementary streams consisting of individually accessible chunks called access units are forwarded to the compression layer where the streams describing structure of MPEG-7 data (Schema streams), as well as the streams describing the content (partial or full description streams) are decoded. Note that descriptions may be represented in the textual form or the binary form, Binary format for Metadatam (BiM). Due to user interaction with the application, request for multimedia media streams may be generated and is fed back via delivery layer to transmission/storage. The actual mechanism of providing the multimedia content is considered to be part of a complete application and thus is outside the scope of the MPEG-7 standard.

Part 2), MPEG-7 DDL, is a standardized language for defining new DSs and Ds, as well as extending or modifying existing DSs and Ds. MPEG-7 DDL is derived by extension of XML Schema, which is developed by W3C consortium using eXtensible Markup Language (XML) as the basis. While the XML Schema has many of the capabilities needed by MPEG-7, it had to be extended to address other requirements specific to MPEG-7. The resulting language satisfies the following key requirements necessary for MPEG-7:

- datatype definition;
- D and description scheme declaration;
- attribute declaration;
- typed reference;
- content model;
- inheritance/subclassing mechanism;
- abstract D and description scheme;
- description scheme inclusion.

Part 3), MPEG-7 Visual, specifies a set of standardized visual Ds and DSs. Visual Ds mainly address specific features such as color, texture, shape and motion. Visual Ds often require other low-level Ds or support elements such as structure (grid layout, spatial coordinates), viewpoint (multiple view), localization (region locator), and temporal (time series, temporal interpolation). A number of Ds for each feature are standardized, such as:

- color Ds, such as Color Space, Color Quantization, Dominant Color, Scalable Color, Color Layout, Color Structure, and Group of Picture Color;

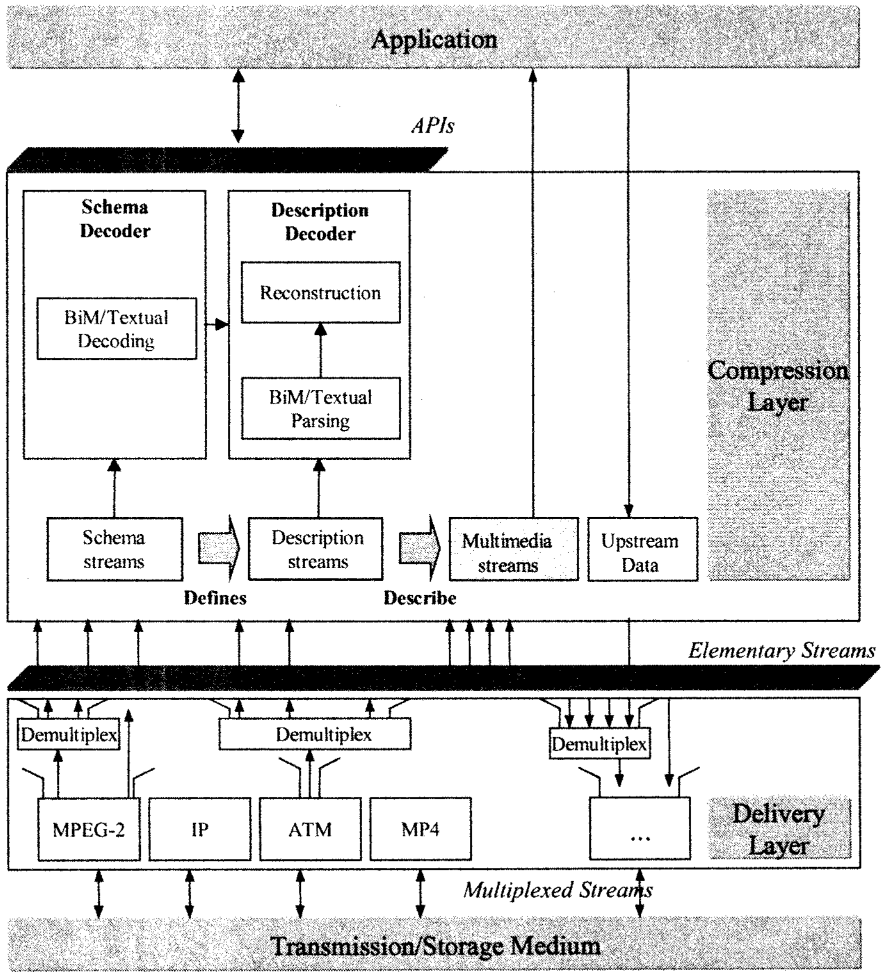

Fig. 3. Architecture of MPEG-7 terminal.

- texture Ds, such as Homogeneous Texture, Texture Browsing, and Edge histogram;
- shape Ds, such as Region Shape, Contour Shape, and Shape 3D;
- motion Ds, such as Camera Motion, Motion Trajectory, Parametric Motion, and Motion Activity;
- other Ds, such as Face Recognition.

In addition, a description scheme for localization of information called SpatioTemporal locator is composed of other DSs, such as FigureTrajectory or ParameterTrajectory, is also defined.

Part 4), MPEG-7 Audio, specifies a set of standardized audio Ds and DSs. MPEG-7 audio Ds address four classes of audio signals: pure music, pure speech, pure sound effects, and arbitrary soundtracks. Audio Ds and DSs may address audio features such as silence, spoken content, timbre, sound effects, melody, etc. Audio Ds often require other low-level D categories such as scalable series (ScalableSeries, SeriesofScalarType, etc.) and Audio Description framework (AudioSampledType, AudioWaveformEnvelopeType, etc). Examples of standardized Ds for various audio features are as follows:

- silence Ds, such as SilenceType;
- spoken content (representation of output of automatic speech recognition) Ds, such as SpokenContentSpeakerType, SpokenContentExtractionInfoType, SpokenContentConfusionInfoType, spokencontentLinkType;
- timbre (perceptual features of instrument sounds) Ds, such as InstrumentTimbreType, HarmonicInstrumentTimbreType, PercussiveInstrumentTimbreType;
- sound effects Ds, such as AudioSpectrumBasisType, SoundEffectFeaturesType;
- melody contour Ds, such as ContourType, MeterType, BeatType.

A number of DSs, such as that for Spoken content, Sound effects, Melody contour, and Melody that utilize the aforementioned Ds have also been defined.

Part 5), MPEG-7 Multimedia Description Schemes (MDS), specifies a high-level framework that allows generic description of all kinds of multimedia including audio, visual, and textual data. This is in contrast with the specific descriptions addressed by the audio or visual parts mentioned above. Fig. 4 shows an overview of levels and relationship between levels in MDS hierarchy. The lowest level, called the *basic elements*, consists of datatypes, mathematical structures, linking and media localization tools, and elementary DSs. The next level, called the *content management & content description*, builds on the lowest level. It describes the content from several viewpoints: creation and production, media, usage, structural aspects, and conceptual as-

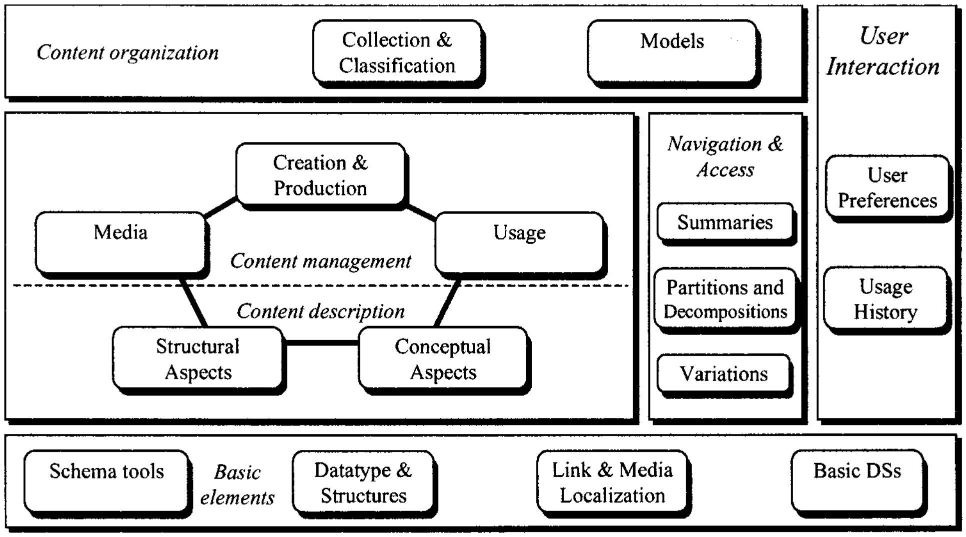

Fig. 4. Overview of MPEG-7 MDS.

pects. The first three elements address primarily information related to the management of the content (*content management*), while the last two are devoted to the description of perceivable information (*content description*).

Beside the direct description of the content provided by these five sets of elements, tools are also defined for *navigation and access*. Summary, Variation and Decomposition elements, allow adaptation of different multimedia presentations to the capabilities of the client terminals, network conditions and user preferences. Some tools are defined for defining user preferences and use history for enhancing the *user interaction* experience. The last set of tools address the *organization* of content by collections and classification, and by use of models.

Part 6), MPEG-7 Reference Software, aims to provide a reference implementation of the relevant parts of the MPEG-7 Standard, and is known as experimentation software (XM). Description of the software, environment, and copyright constraints can be found in [5], [6]. Although some software for extracting Ds is also included, the focus is on creating bitstreams of Ds and DSs with normative syntax, rather than the performance of the tools. Currently, it includes components in four categories: DDL parser and DDL validation parser, visual Ds, audio Ds, and multimedia DSs.

Part 7), MPEG-7 Conformance, aims to provide guidelines and procedures for testing the conformance of MPEG-7 implementations, and has only recently been started.

IV. Relationship with Other Standards

While MPEG-7 focuses on description of content contained in generic audio–visual information, there are other metadata standards developed or being developed by other communities for various types of information in various domains. Some examples include SMPTE [9], EBU [8], TV-Anytime [10], DIG-35 [12], Dublin Core,[1] and OCLC/RLG [11]. We briefly compare these standards according to three criteria (as those used in [11]): 1) their rationales and objectives; 2) their underlying framework; and 3) the elements constituting the concepts or higher-level structures in these standards.

[1]Dublin Core Metadata Initiative. [Online]. Available: http://purl.org/DC

As described in the previous paragraph, MPEG-7 aims at a generic, comprehensive solution for describing audio–visual information from various domains. Although there is ongoing discussion about "profiling" to define subsets of tools for specific applications, most parts of MPEG-7 are application-, technology-, and platform-independent. It is intended to facilitate and support push-type (filter), as well as pull-type (retrieval), applications. The process of development is based on identification of fundamental concepts through conceptual modeling, definition of underlying Ds, structural construction of DSs, adoption of XML Schema as the definition language, and specification of the systems part such as formats for file storage and transport. The final framework enables highly structural, detailed descriptions of audio–visual information at different granularity levels (region, image, video segment, collection) and in different areas (description, management, organization, navigation, and user interaction).

A. SMPTE/EBU

SMPTE/EBU- related standards [8], [9], and [13] focus on metadata about the production, broadcasting, and distribution of different types of essence (video, audio, and data, in various forms). Although it is also intended to support content distribution over Internet, the main focus is on the distribution over broadcast channels. The current SMPTE standard is based on a "dictionary" model, which consists of about 1000 elements. The approach supports some simple grouping mechanisms including "like things" (e.g., similar ISO identifiers), "type node" (e.g., different representations for the same element), and "set" (e.g., a poem's title, abstract and its poet's name). The relationships between elements are not as detailed and complex as those defined in MPEG-7.

B. TV-Anytime

TV-Anytime [7], [10] focuses on the specifications enabling enhanced audio–visual services based on the consumer-side local storage, the so-called Personal Digital Recorder. It includes three parts: content referencing, metadata, and rights management. Like MPEG-7, the TV-Anytime Forum has adopted XML as the common representation format for metadata. Its metadata is based on the Description Scheme framework and is specified in MPEG-7 DDL. At the same time, TV-Anytime is also defining a dictionary to provide definitions of elements and attributes used in various schemes. There have been some MPEG-7 tools adopted in TV-Anytime, such as Hierarchical Summary DS and User Preferences DS. There is also current harmonization effort between MPEG-7 and TV-Anytime to address the Electronic Program Guide (EPG) DS. It should be noted that TV-Anytime will include tools from MPEG-7, but also non-MPEG-7 tools (such as descriptions of channels and services).

C. DIG-35

DIG-35 [12] aims at interoperable specifications for a standard set of metadata for digital images. Driven mostly by the digital photography industry, the standard focuses on descriptions of image parameters, creation information, content, history, and rights management information. Procedural information, such as processing, printing, and ordering information, is not included. The framework is also based on XML and the constituent elements have a high extent of correspondence with the elements defined in MPEG-7, except that audio and video related information is not covered in DIG-35.

D. Dublin Core

The Dublin Core Element Set[2] provides a metadata model, including 15 basic elements for discovery of electronic resources. It is mainly used by libraries, museums, government agencies, and commercial organizations. The consensus-based elements include very simple elements such as title, creator, subject, description, format, rights, and others. However, compared to the detailed structures defined in MPEG-7, they can be accessed quite efficiently. Rather than using a formal process to map the Dublin Core elements to corresponding MPEG-7 elements, there have been some efforts in integrating elements from both Dublin Core and MPEG-7 in the same description scheme by using the XML Namespaces approach [14].

E. OCLC/RLG

Finally, the Digital Object Preservation Metadata is being addressed by the OCLC/RLG group [11], with the primary objective in developing a comprehensive preservation metadata framework for digital preservation activities mainly in libraries or academic institutions. The digital object could be in any form of recorded knowledge, e.g., the scanned files of printed documents, digital audio clips, digital images, an online HTML files, or others. By consensus, the group has adopted the Open Archival Information System (OAIS) reference model as a foundation. OAIS provides a conceptual framework to model the functional components and processes common in archival activities as well as an information model for defining the metadata components. Four basic types of information are addressed: content information (i.e., the archived object and information about its representation like file format), preservation description information (like reference identifiers, history of preservation processes, authentication information), packaging information (for binding the data object and its associated metadata), and descriptive information (for aiding the search and retrieval tools). Currently, several digital preservation initiatives have adopted the OAIS model, although issues of appropriate levels of granularity of defining and applying metadata are still open issues. It should be noted that the XML-based approach is also being considered as a potential representation framework.

In comparing MPEG-7 with other standards, the following observations can be made. First, MPEG-7 and other standards share the same objective in developing an interoperable framework of metadata description for general types of objects. While MPEG-7 aims at supporting a broad range of applications, other standards focus more on specific application domains or activities. Therefore, it will be a worthwhile effort to harmonize or synchronize MPEG-7 with other domain-specific approaches. Second, MPEG-7 currently uses a Schema approach in defining elements, attributes and structures for different concepts, while other standards like SMPTE use a dictionary approach. As a result, MPEG-7 supports a more detailed, structured description of audio–visual information at multiple levels of granularity. Integration or mapping mechanisms for ensuring the interoperability between standards with different frameworks will be an important task. Currently, a group in MPEG-7 in actively working on this integration task [13].

V. OPEN ISSUES AND DISCUSSION

As the standard is being completed, MPEG-7 is faced with several evolving and important challenges.

A. Conformance and Profiling

Definition of interoperability and conformance testing is an important issue that is being actively discussed [16]. Audio–visual descriptions are said to be MPEG-7 conformant if they follow the defined syntax and semantics, which will be unambiguously understood by MPEG-7 compliant terminals and systems. At the systems level, binary and textual representations of conformant descriptions should be parsed correctly by the conformant textural or binary decoders. At the semantic level, the meanings of the elements and structures used in the descriptions should be understood by the description consuming systems in the context of MPEG-7 audio, visual, or MDS. However, given a set of audio–visual descriptions, it will be difficult to test the conformance of semantics of the Ds or DSs used. For example, it is difficult to verify whether the values of the color D faithfully and accurately represent the color feature of the image. For DSs, it is difficult to verify the reliability of the relationships between different elements specified in an instantiation.

In addition, the behaviors of the description consuming systems will not be normative, and will depend on the specific applications (e.g., search, filtering, or adapation). The introduc-

[2]Dublin Core Metadata Initiative. [Online]. Available: http://purl.org/DC

tion of the profiling concept may be necessary and important for MPEG-7 to define unambiguous conformance procedure and interoperability. Profiles and levels have been used in other audio–visual standards (e.g., MPEG-2 and MPEG-4) to define a subset of tools with a restricted level of complexity that are useful for specific applications. For MPEG-7, it is clear that different applications will need different subsets of tools. Some specialized applications may also need restrictions or extensions of the MPEG-7 tools. How MPEG-7 ought to support extensions and specialization in specific domains and at the same time provides unambiguous conformance validation is a critical issue and requires a nontrivial, timely effort.

B. Intellectual Property Management and Protection

In the context of MPEG-7, the Intellectual Property Management and Protection (IPMP) issues involve management and protection of descriptions (i.e., instantiations of the Ds and DSs) and tools (i.e., new Ds and DSs defined in DDL), apart from the audio–visual content. Currently, an active effort [15] is being undertaken in developing mechanisms to: 1) manage intellectual properties; 2) prevent unauthorized access, use, and modification; and 3) support authentication of descriptions and tools. Techniques developed in W3C for signing XML documents and solutions used for MPEG-4 IPMP support are being considered. An important issue is to allow for flexible IPMP support at appropriate levels of granularity (e.g., descriptions, elements, or attributes) and in dynamic environments (e.g., dynamic transmission of descriptions for streamed content). Another critical issue is the robustness of the IPMP solutions through the process of acceptable transformations (e.g., canonicalization of XML documents). It is desirable that the IPMP data can survive acceptable transformations so that end-to-end protection or authentication can be achieved.

C. Impact

Finally, with the tools and technologies of MPEG-7 ready for implementation and deployment, it is important to find the path for MPEG-7 to enter practical application domains and accomplish the claimed objectives. A number of groups from industry have shown active interest in promoting the awareness, trials, and adoption of the MPEG-7 standard.[3] In addition, the active efforts in harmonizing the MPEG-7 standard with other metadata standards in emerging applications (e.g., TV-Anytime and SMPTE/EBU applications) will facilitate and expedite the adoption of MPEG-7 in practical domains.

ACKNOWLEDGMENT

The ISO MPEG-7 standard results from contributions and collaborative efforts of a large number of experts. This paper contains authors' personal understanding and views of the standard. For fully updated and accurate information, readers are referred to the latest official documents and related publications. The authors appreciate the valuable comments from anonymous reviewers and A. Belen Benitez, who is a co-editor of MPEG-7 Part 5.

[3][Online]. Available: http://www.mpeg-7.com/

REFERENCES

[1] *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 685–772, June 2001.

[2] *MPEG-7 Requirements Document V.13*, ISO/IEC JTC1/SC29/WG11/N3933, Jan. 2001.

[3] *Introduction to MPEG-7 (V 2.0)*, ISO/IEC JTC1/SC29/WG11/N3751, July 2000.

[4] *MPEG-7: Context, Objectives and Technical Roadmaps (V.12)*, ISO/IEC JTC1/SC29/WG11/N2861, July 1999.

[5] *ISO/IEC CD 15938-(1–6) Information Technology— Multimedia Content Description Interface Part 1–6*, ISO/IEC JTC1/SC29/WG11/N3701-N3706, Oct. 2000.

[6] A. Yamada, M. Pickering, S. Jeannin, L. Cieplinski, J.-R. Ohm, and M. Kim, Eds., (Oct. 2000) *MPEG-7 Visual part of eXperimentation Model Version 8.0*, ISO/IEC JTC1/SC29/WG11/N3673. [Online]. Available: http://www.cselt.it/mpeg/public/

[7] *MPEG-7 Report on The AHG on the TV-Anytime Application*, ISO/IEC JTC1/SC29/WG11/M6485, Oct. 2000.

[8] European Broadcasting Union. *PMC Project P/META (Metadata Exchange Standards)*. [Online]. Available: http://www.ebu.ch/pmc_meta.html

[9] *Proposed SMPTE 335M Specification for Metadata Dictionary Structure*, Jan. 2000.

[10] The TV-Anytime Forum, Specification Series: S-3 on Metadata Document: TV049 (2000, Sept.). [Online]. Available: ftp://tva:tva@ftp.bbc.co.uk/pub/Plenary/

[11] "Preservation Metadata for Digital Objects: A Review of the State of the Art ," OCLC/RLG Working Group on Preservation Metadata, Jan. 2001.

[12] (2000, Mar.) DIG35 Metadata for Digital Images, Working Draft 1.0. [Online]. Available: www.digitalimaging.com

[13] *MPEG-7: AHG on MPEG-7 Integration with SMPTE and EBU, AHG Report*, ISO/IEC JTC1/SC29/WG11/M6288, July 2000.

[14] *MPEG-7: Proposal for the Integration of Dublin Core and MPEG-7*, ISO/IEC JTC1/SC29/WG11/M6500, Oct. 2000.

[15] *MPEG-7 IPMP—First Technical Ideas and Architecture*, ISO/IEC JTC1/SC29/WG11/N3876, Jan. 2001.

[16] *MPEG-7 Interoperability, Conformance Testing and Profiling, Version 2*, ISO/IEC JTC1/SC29/WG11/N4039, Mar. 2001.

Shih-Fu Chang (S'89–M'93–SM'01) is an Associate Professor of Electrical Engineering at Columbia University, New York, where he currently leads the ADVENT university–industry consortium and the digital video and multimedia group. The group conducts research in searching, filtering, manipulation, transmission, and security of digital video and multimedia information, in addition to having played an active role in developing the ISO MPEG-7 Standard. He has actively participated in several large-scale cross-disciplinary projects, including Columbia's Digital Library Project, Digital News System, and Art Image Education for K–12 Schools. He and his group have developed several well-known systems and algorithms for video searching, like VideoQ, WebSEEk, and compressed video manipulation, like WebClip. He has also served as a Consultant in several new media technology companies.

Prof. Chang is currently a Distinguished Lecturer of the IEEE Circuits and Systems Society in the area of multimedia systems and technology. He was a General Co-Chair of the ACM 8th Multimedia Conference 2000 and an Associate Editor for several journals. He was awarded the Navy ONR Young Investigator Award, a CAREER Award from the the National Science Foundation, a Faculty Development Award from IBM, a Service Recognition Award from ACM, and has received three Best Paper Awards from IEEE, ACM, and SPIE.

Thomas Sikora (M'93–SM'96) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Bremen University, Germany, in 1985 and 1989, respectively.

In 1990, he joined Siemens Ltd., Melbourne, Australia, and Monash University, Melbourne, Australia, as a Project Leader, responsible for video compression research activities in the Australian "Universal Broadband Video Codec" consortium. Between 1994 and 1998, he directed video-coding research activities at Heinrich–Hertz-Institute (HHI), Berlin, Germany. He is currently Head of the Interactive Media—Human Factors Department at HHI, in addition to being an Adjunct Professor in the Electrical Engineering Department, Technical University Berlin, Berlin, Germany. He is co-founder of $2SK$ Media Technologies and Perspective Technologies GmbH, two recent Berlin-based start-up companies involved in audio and video signal processing, compression technology and 3-D visualization. He has been involved in ITU and ISO standardization activities, as well as in several European research activities for a number of years. As the chairman of the ISO-MPEG video group, he is responsible for the development and standardization of the MPEG video-coding algorithms. Between 1995–1998, he also served as the chairman of the European COST $211^{\rm ter}$ video compression research group. He is frequently an industry consultant on issues related to interactive digital audio and video.

Dr. Sikora was the recipient of the 1996 German ITG Award (German Society for Information Technology). He is an Associate Editor for a number of international journals, including the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *IEEE Signal Processing Magazine*, the *EURASIP Signal Processing Journal*, and the *EURASIP Signal Processing: Image Communication Journal*. He is a member of ITG.

Atul Puri (S'86–M'87) received the M.S. degree from the City College of New York in 1982, and the Ph.D. degree a from the City University of New York in 1988, both in electrical engineering.

He was a Consultant in the Visual Communications Research Department, Bell Labs, Holmdel, NJ, which he joined as a member of Technical Staff in 1988. Since 1996, he has been a Technology Consultant in image processing research, AT&T Labs, Middletown, NJ, having previously been a Principal Technical Staff Member. He has been involved in video-coding research for videoconferencing, video on digital storage media, HDTV, and stereoscopic TV applications. His current research interests are in the areas of flexible video and multimedia systems and services for the Web/Internet. He has represented AT&T at ISO MPEG audio–visual coding standards, such as MPEG-1, MPEG-2, and MPEG-4. He is a Technical Editor of the Video part of the MPEG-4 standard and is also involved in the MPEG-7 multimedia content description standard, currently in progress. He holds over 24 patents and has published over 40 technical papers. He is a coauthor of the book *Digital Video: An Introduction to MPEG-2* (London, U.K.: Chapman & Hall, 1997) and co-editor of the book *Multimedia Systems, Standards and Networks* (New York: Dekker, 2000).

Dr. Puri is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He has received a number of awards from AT&T, including the Exceptional Contribution Award, the Individual Performance Award, and the Standards Recognition Award. He is also a recipient of the ISO Prize. He is a member of IEEE Communications Society and the IEEE Signal Processing Society.